

WHAT IS CLAIMED

sub
R1

1 1. A method for searching a data repository managed by a content provider
2 to gather indexable metadata on content at addresses locations at the data repository,
3 comprising:
4 accessing settings capable of being customized by the content provider, wherein
5 the customized settings provide instructions on how to search the content provider's data
6 repository;
7 accessing content pages at the content provider's data repository;
8 accessing the content of content pages at the content provider's data repository in
9 accordance with instructions included in the accessed customized settings; and
10 generating metadata from accessed content pages to add to an index of metadata
11 for accessed addressable locations at the data repository.

1 2. The method of claim 1, wherein the customized settings include
2 parameters and access methods unique to an arrangement of content in the content
3 provider's data repository.

1 3. The method of claim 1, wherein the accessed customizable settings
2 provide addressable locations at the content provider's data repository provided by the
3 content provider, wherein accessing the content pages includes accessing the content
4 pages at the provided addressable locations, wherein metadata is generated for the
5 accessed content pages.

1 4. The method of claim 3, wherein the addressable locations comprise
2 uniform resource locator (URL) addresses.

1 5. The method of claim 3, wherein the accessed customizable settings
2 provide query terms for at least one included addressable location, further comprising:

004230-94254960

3 for each provided addressable location for which there are query terms, using the
4 provided query terms at the provided addressable location to obtain query results; and
5 generating metadata from the obtained query results to add to the index of
6 metadata for accessed addressable locations at the data repository.

1 6. The method of claim 5, wherein the accessed customizable settings further
2 provide qualifiers for at least one search term, further comprising:

3 for each query term having at least one qualifier, determining whether the query
4 results for the query term satisfy each qualifier for the query term, wherein the metadata
5 for the query result is generated if the query result satisfies each qualifier for the query
6 term that generated the query result; and
7 performing a non-qualifying action for each query result that does not satisfy each
8 qualifier.

1 7. The method of claim 6, wherein the non-qualifying action comprises not
2 including metadata for the query result in the index.

1 8. The method of claim 3, wherein the accessed customizable settings further
2 provide a password for at least one provided addressable location, further comprising:
3 using the provided password to access the content page at the indicated
4 addressable location for which the password is provided.

1 9. The method of claim 1, wherein the accessed customizable settings further
2 indicate a recursive search setting indicating whether to search hypertext links to linked
3 addressable locations included in the accessed content page, further comprising:
4 accessing a content page at each linked addressable location included if the
5 recursive search setting indicates to recursively search linked addressable locations,
6 wherein metadata is generated for each content page recursively accessed at the linked
7 addressable locations in the accessed content page.

004230 9825430

1 10. The method of claim 9, wherein the accessed customizable settings further
2 provide prohibited addressable locations at the data repository, wherein metadata is not
3 generated for each content page at a linked addressable location that is one indicated
4 prohibited address location.

1 11. The method of claim 1, wherein the accessed customizable settings further
2 indicate validation checking programs, further comprising:
3 executing each validation checking program indicated in the accessed
4 customizable settings against each accessed content page;
5 generating a validation output result with the validation checking program for
6 each accessed content page with each validation checking program describing
7 characteristics of the content page;
8 generating metadata from the validation output result to add to the index of
9 metadata for accessed addressable locations at the data repository.

1 12. The method of claim 11, wherein the accessed customizable settings
2 further indicate at least one parameter to use with at least one validation checking
3 program, further comprising:
4 using the at least one parameter when executing the validation checking program,
5 wherein the validation output result further indicates characteristics of the content page
6 related to the at least one parameter used with the validation checking program.

1 13. The method of claim 11, wherein the accessed customizable settings
2 further indicate at least one qualifier to use with at least one validation checking program,
3 further comprising:
4 determining whether the validation output result satisfies the at least one qualifier
5 provided with the validation checking program producing the output result, wherein
6 metadata for the output result is included in the index if the output result satisfies the
7 qualifier.

007200360US1

1

1 14. The method of claim 13, wherein metadata for the content page at the
2 addressable location is not included in the index if the validation output result does not
3 satisfy the qualifier.

1 15. The method of claim 1, further comprising:
2 determining a format of the accessed content page;
3 selecting one of a plurality parsers capable of parsing the determined format; and
4 parsing the content page using the selected parser, wherein the metadata to add to
5 the index is generated from the parsed content page.

1 16. The method of claim 1, further comprising:
2 determining a parser capable of parsing an embedded file referenced in the content
3 page;
4 parsing the content of the referenced embedded file; and
5 generating metadata for the parsed content of the embedded file to add to the
6 index.

1 17. The method of claim 14, wherein the embedded file is encoded in a
2 multimedia format.

1 18. The method of claim 1, further comprising:
2 distributing a collection tool to content providers capable of accessing and
3 generating metadata for content provider data repositories using the accessed
4 customizable settings; and
5 collecting metadata data gathered from multiple content providers using the
6 collection tool to gather metadata on their data repositories;

1 19. The method of claim 18, further comprising commercializing the collected
2 metadata.

004230:9363490

81

1 20. The method of claim 18, further comprising:
2 receiving an electronic subscription from content providers to use the collection
3 tool and provide metadata.

1 21. A system for searching a data repository managed by a content provider to
2 gather indexable metadata on content at addresses locations at the data repository,
3 comprising:

4 means for accessing settings capable of being customized by the content provider,
5 wherein the customized settings provide instructions on how to search the content
6 provider's data repository;

7 means for accessing content pages at the content provider's data repository;

8 means for accessing the content of content pages at the content provider's data
9 repository in accordance with instructions included in the accessed customized settings;
10 and

11 means for generating metadata from accessed content pages to add to an index of
12 metadata for accessed addressable locations at the data repository.

1 22. The system of claim 21, wherein the customized settings include
2 parameters and access methods unique to an arrangement of content in the content
3 provider's data repository.

1 23. The system of claim 22, wherein the accessed customizable settings
2 provide addressable locations at the content provider's data repository provided by the
3 content provider, wherein the means for accessing the content pages includes accessing
4 the content pages at the provided addressable locations, wherein metadata is generated for
5 the accessed content pages.

1 24. The system of claim 23, wherein the addressable locations comprise
2 uniform resource locator (URL) addresses.

0042303354960

21

1 25. The system of claim 23, wherein the accessed customizable settings
2 provide query terms for at least one included addressable location, further comprising:
3 means for using the provided query terms at the provided addressable location to
4 obtain query results for each provided addressable location for which there are query
5 terms; and
6 means for generating metadata from the obtained query results to add to the index
7 of metadata for accessed addressable locations at the data repository.

1 26. The system of claim 25, wherein the accessed customizable settings
2 further provide qualifiers for at least one search term, further comprising:
3 means for determining whether the query results for the query term satisfy each
4 qualifier for the query term for each query term having at least one qualifier, wherein the
5 metadata for the query result is generated if the query result satisfies each qualifier for the
6 query term that generated the query result; and
7 means for performing a non-qualifying action for each query result that does not
8 satisfy each qualifier.

1 27. The system of claim 26, wherein the non-qualifying action comprises not
2 including metadata for the query result in the index.

1 28. The system of claim 22, wherein the accessed customizable settings
2 further provide a password for at least one provided addressable location, further
3 comprising:
4 means for using the provided password to access the content page at the indicated
5 addressable location for which the password is provided.

1 29. The system of claim 23, wherein the accessed customizable settings
2 further indicate a recursive search setting indicating whether to search hypertext links to
3 linked addressable locations included in the accessed content page, further comprising:

007230" 33E5450

81

4 means for accessing a content page at each linked addressable location included if
5 the recursive search setting indicates to recursively search linked addressable locations,
6 wherein metadata is generated for each content page recursively accessed at the linked
7 addressable locations in the accessed content page.

1 30. The system of claim 29, wherein the accessed customizable settings
2 further provide prohibited addressable locations at the data repository, wherein metadata
3 is not generated for each content page at a linked addressable location that is one
4 indicated prohibited address location.

1 31. The system of claim 21, wherein the accessed customizable settings
2 further indicate validation checking programs, further comprising:
3 means for executing each validation checking program indicated in the accessed
4 customizable settings against each accessed content page;
5 means for generating a validation output result with the validation checking
6 program for each accessed content page with each validation checking program
7 describing characteristics of the content page; and
8 means for generating metadata from the validation output result to add to the
9 index of metadata for accessed addressable locations at the data repository.

1 32. The system of claim 31, wherein the accessed customizable settings
2 further indicate at least one parameter to use with at least one validation checking
3 program, further comprising:
4 means for using the at least one parameter when executing the validation checking
5 program, wherein the validation output result further indicates characteristics of the
6 content page related to the at least one parameter used with the validation checking
7 program.

004230" 93654960

A1

1 33. The system of claim 31, wherein the accessed customizable settings
2 further indicate at least one qualifier to use with at least one validation checking program,
3 further comprising:

4 means for determining whether the validation output result satisfies the at least
5 one qualifier provided with the validation checking program producing the output result,
6 wherein metadata for the output result is included in the index if the output result satisfies
7 the qualifier.

1 34. The system of claim 33, wherein metadata for the content page at the
2 addressable location is not included in the index if the validation output result does not
3 satisfy the qualifier.

1 35. The system of claim 21, further comprising:
2 means for determining a format of the accessed content page;
3 means for selecting one of a plurality parsers capable of parsing the determined
4 format; and
5 means for parsing the content page using the selected parser, wherein the metadata
6 to add to the index is generated from the parsed content page.

1 36. The system of claim 21, further comprising:
2 means for determining a parser capable of parsing an embedded file referenced in
3 the content page;
4 means for parsing the content of the referenced embedded file; and
5 means for generating metadata for the parsed content of the embedded file to add
6 to the index.

1 37. The system of claim 36, wherein the embedded file is encoded in a
2 multimedia format.

0042200360US1

R1

1 38. The system of claim 22, further comprising:
2 means for distributing a collection tool to content providers capable of accessing
3 and generating metadata for content provider data repositories using the accessed
4 customizable settings; and
5 means for collecting metadata data gathered from multiple content providers using
6 the collection tool to gather metadata on their data repositories;

1 39. A program for searching a data repository managed by a content provider
2 to gather indexable metadata on content at addresses locations at the data repository,
3 wherein the program comprises code implemented in a computer readable medium
4 capable of causing a computer to perform:
5 accessing settings capable of being customized by the content provider, wherein
6 the customized settings provide instructions on how to search the content provider's data
7 repository;
8 accessing content pages at the content provider's data repository;
9 accessing the content of content pages at the content provider's data repository in
10 accordance with instructions included in the accessed customized settings; and
11 generating metadata from accessed content pages to add to an index of metadata
12 for accessed addressable locations at the data repository.

1 40. The method of claim 39, wherein the customized settings include
2 parameters and access methods unique to an arrangement of content in the content
3 provider's data repository.

1 41. The program of claim 39, wherein the accessed customizable settings
2 provide addressable locations at the content provider's data repository provided by the
3 content provider, wherein accessing the content pages includes accessing the content
4 pages at the provided addressable locations, wherein metadata is generated for the
5 accessed content pages.

007200360US1

1 42. The program of claim 41, wherein the addressable locations comprise
2 uniform resource locator (URL) addresses.

1 43. The program of claim 39, wherein the accessed customizable settings
2 provide query terms for at least one included addressable location, wherein the program is
3 further capable of causing the computer to perform:
4 for each provided addressable location for which there are query terms, using the
5 provided query terms at the provided addressable location to obtain query results; and
6 generating metadata from the obtained query results to add to the index of
7 metadata for accessed addressable locations at the data repository.

1 44. The program of claim 43, wherein the accessed customizable settings
2 further provide qualifiers for at least one search term, wherein the program is further
3 capable of causing the computer to perform:
4 for each query term having at least one qualifier, determining whether the query
5 results for the query term satisfy each qualifier for the query term, wherein the metadata
6 for the query result is generated if the query result satisfies each qualifier for the query
7 term that generated the query result; and
8 performing a non-qualifying action for each query result that does not satisfy each
9 qualifier.

1 45. The program of claim 44, wherein the non-qualifying action comprises not
2 including metadata for the query result in the index.

1 46 The program of claim 39, wherein the accessed customizable settings
2 further provide a password for at least one provided addressable location, wherein the
3 program is further capable of causing the computer to perform:
4 using the provided password to access the content page at the indicated
5 addressable location for which the password is provided.

0042303365450

1 47. The program of claim 39, wherein the accessed customizable settings
2 further indicate a recursive search setting indicating whether to search hypertext links to
3 linked addressable locations included in the accessed content page, wherein the program
4 is further capable of causing the computer to perform:

5 accessing a content page at each linked addressable location included if the
6 recursive search setting indicates to recursively search linked addressable locations,
7 wherein metadata is generated for each content page recursively accessed at the linked
8 addressable locations in the accessed content page.

1 48. The program of claim 47, wherein the accessed customizable settings
2 further provide prohibited addressable locations at the data repository, wherein metadata
3 is not generated for each content page at a linked addressable location that is one
4 indicated prohibited address location.

1 49. The program of claim 39, wherein the accessed customizable settings
2 further indicate validation checking programs, wherein the program is further capable of
3 causing the computer to perform:
4 executing each validation checking program indicated in the accessed
5 customizable settings against each accessed content page;
6 generating a validation output result with the validation checking program for
7 each accessed content page with each validation checking program describing
8 characteristics of the content page;
9 generating metadata from the validation output result to add to the index of
10 metadata for accessed addressable locations at the data repository.

1 50. The program of claim 49, wherein the accessed customizable settings
2 further indicate at least one parameter to use with at least one validation checking
3 program, wherein the program is further capable of causing the computer to perform:

004456 03E54560

4 using the at least one parameter when executing the validation checking program,
5 wherein the validation output result further indicates characteristics of the content page
6 related to the at least one parameter used with the validation checking program.

1 51. The program of claim 49, wherein the accessed customizable settings
2 further indicate at least one qualifier to use with at least one validation checking program,
3 wherein the program is further capable of causing the computer to perform:
4 determining whether the validation output result satisfies the at least one qualifier
5 provided with the validation checking program producing the output result, wherein
6 metadata for the output result is included in the index if the output result satisfies the
7 qualifier.

1 52. The program of claim 51, wherein metadata for the content page at the
2 addressable location is not included in the index if the validation output result does not
3 satisfy the qualifier.

1 53. The program of claim 39, wherein the program is further capable of
2 causing the computer to perform:
3 determining a format of the accessed content page;
4 selecting one of a plurality parsers capable of parsing the determined format; and
5 parsing the content page using the selected parser, wherein the metadata to add to
6 the index is generated from the parsed content page.

1 54. The program of claim 39, wherein the program is further capable of
2 causing the computer to perform:
3 determining a parser capable of parsing an embedded file referenced in the content
4 page;
5 parsing the content of the referenced embedded file; and

007200360US1

A1

6 generating metadata for the parsed content of the embedded file to add to the
7 index.

1 55. The program of claim 54, wherein the embedded file is encoded in a
2 multimedia format.

1 56. The program of claim 39, further comprising:
2 distributing the program to content providers capable of accessing and generating
3 metadata for content provider data repositories using the accessed customizable settings;
4 and
5 collecting metadata data gathered from multiple content providers using the
6 collection tool to gather metadata on their data repositories;

1 57. The program of claim 56, further comprising:
2 receiving an electronic subscription from content providers to use the program to
3 gather and provide metadata.

00720032